Investigations into ATSR-1 data artefacts

	Name	Organisation	Signature and date
Prepared	Andy Smith	RAL	Digitally signed by Andy Smith DN: cn=Andy Smith, o=STFC, ou=RAL Space, email=andy.smith@stfc.ac.uk, c=GB Date: 2016.08.05 11:30:33 +01'00'
Checked	Hugh Kelliher	Space ConneXions Limited	
Approved	n/a	RAL	n/a

Science and Technology Facilities Council
Rutherford Appleton Laboratory
Harwell Science and Innovation Campus
Didcot
Oxfordshire OX11 0QX
United Kingdom

Distribution

Name	Organisation
ATSR team	RAL Space
Hugh Kelliher	Space ConneXions Limited
Ruth Wilson	Space ConneXions Limited
Miguel Martinez-Boti	DECC
ATSR QWG	Various

Change Record

Issue	Date	Description
0.1	27-May-2016	New document
0.2	28-Jul-2016	Reviewed by Space ConneXions Limited. Minor updates by AS following review.
1.0	05-Aug-2016	Draft 0.2 checked by Space ConneXions Limited. Accepted all suggested changes ready for issue.

Table of Contents

1 Scope of Document

This Technical Note describes work performed by RAL Space to investigate and deal with data "artefacts" in the ATSR-1 UBT data set. This work is performed as Work Package 4.4 of the proposal described in [AD 1].

2 Terms, Definitions and Abbreviations

2.1 Acronyms

APP	Archive Product Processor (RAL software to process ATSR-1 & 2 UBT to L1B TOA)
ATSR	Along-Track Scanning Radiometer
IDL	Interactive Data Language (data manipulation and graphics software)
LRDAF	Low-Rate Data acquisition Facility (TBC)
QC	Quality Control
SADIST(-2)	Synthesis of ATSR(-2) Data Into Sea surface Temperature (RAL software to process Level 0 ATSR data from tape into UBT products)
SUPPLE	Sadist UBT Processor Linux Environment. Linux-ported version of the SADIST L0 to UBT processor software
TOA	Top of Atmosphere
UBT	Ungridded Brightness Temperature (Level 1 ATSR product, generated at single-scene spatial coverage, 512x512km)

3 Documents

3.1 Applicable Documents

Ref	Title	Document code	Version	Date
AD 1	ATSR Satellite Dataset Supporting Activities, 2014 - 2017	Proposal 2014-07-001 (response to DECC ITT : TRN 829/06/2014)	2	28-Jul-2014
AD 2	SADIST-2 v100 Products	ER-TN-RAL-AT-2164		06-Sep-1995

3.2 Reference Documents

Ref	Title	Document code	Version	Date
RD 1	ATSR-1 artefacts	N/A (internal document)	N/A	04-Mar-2010
RD 2	ATSR-1 artefacts 2	N/A (internal document)	N/A	21-May-2010
RD 3	IDL – Interactive Data Language	http://www.harrisgeospatial.com/docs/using_idl_home.html	8.5	2015

4 Introduction and background

[AD 1] describes the problem and proposed work package as follows.

Scan-line defects are seen in some UBT images. These originate in the real-time down-linked portion of L0 data from some "LRDAF" tapes used in UBT production. It is not possible to correct these defects, but the following approach would improve the situation by flagging affected scan lines with appropriate exception values (similar to the approach used in other problem situations).

The following steps would be required:

1. Identify the UBTs with the problem:
    a. create a utility to identify UBTs with blemishes (this will be based on an as-yet untested algorithm);
    b. test and validate the algorithm/code.
2. If the product-identification algorithm is successful, develop code to:
    a. identify affected scans and flag these as invalid within the affected UBTs;
    b. test and validate the new code.
3. Processing:
    a. process all affected products in the archive;
    b. validate processed data and update the archive.

To allow the new code to be developed rapidly, IDL would be used, throughout, including the actual processing.

Jack Abolins did some initial work on data artefacts and wrote two short background documents in 2010, recording a couple of specific examples. Note there is no comprehensive list of affected products.

Artefacts were originally seen on inspection of ATSR-1 L1B TOA products. Further inspection of corresponding UBT products showed that the artefacts were also present in the UBTs and were not introduced by the APP UBT to L1B processor.

From checks of a small number of UBT products from March 1994 it was concluded that artefacts appeared to be present only in the last 10-15% of the affected tape dumps. "Tape dumps" in this context is taken to mean the down-link from the ERS satellite to the ground station.

Jack Abolins was able to process several example scenes from L0 tape to UBT using tapes from different ground receiving stations, with differing results. It was possible to create "clean" versions of the UBT products where artefacts had been present, by processing from a different Level 0 input file. Jack listed examples that were processed in 2004 and 2005 using the RAL SADIST-2 v356 software. At that time RAL had been routinely processing data from tape to UBT using SADIST software for over a decade.

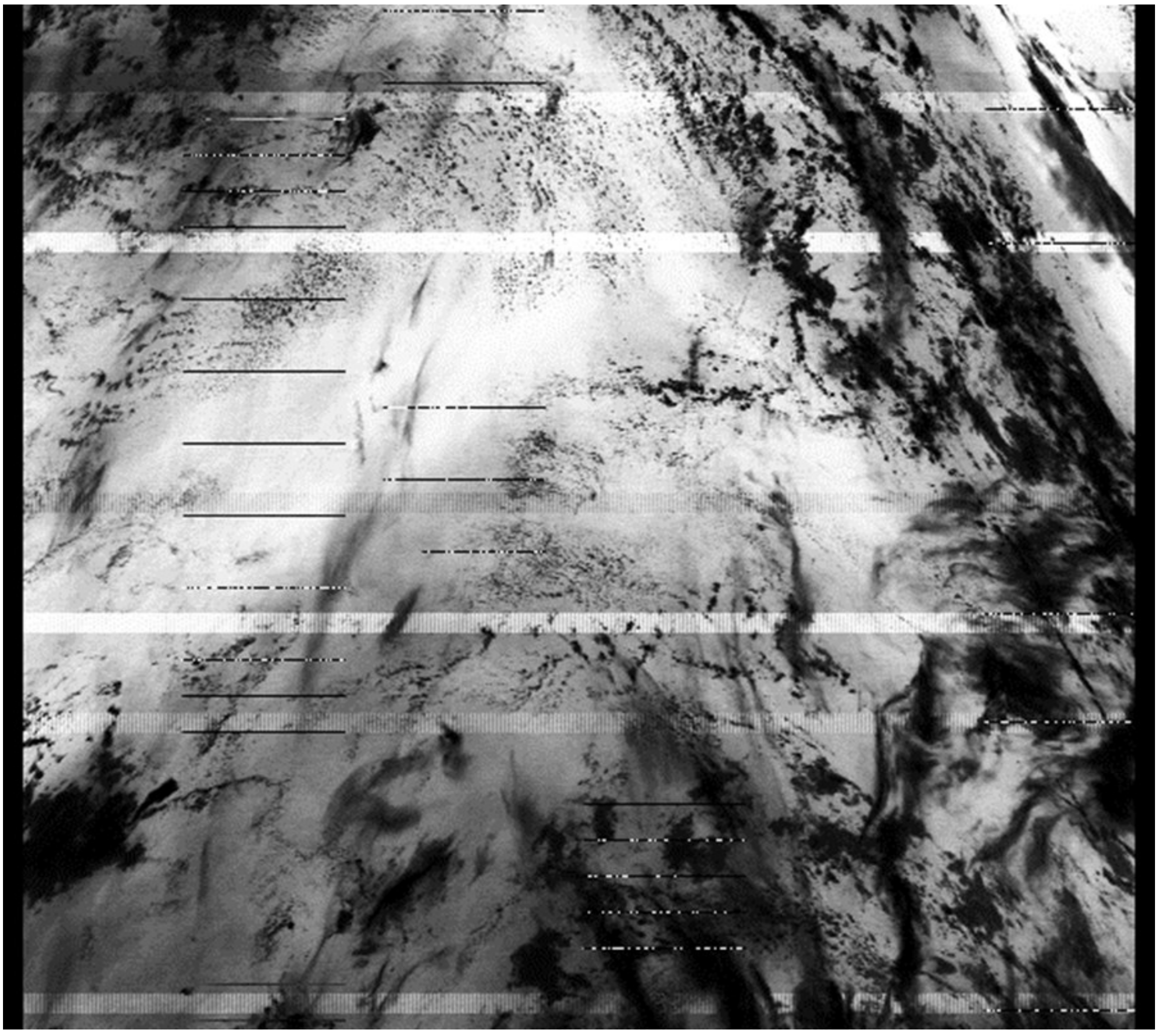

Figure 1: example of a UBT scene from Sept 1991 affected by data artefacts

5 Investigations performed

Figure 1 illustrates 3 of the 4 types of artefact that have been found to date.

1. A “dash” i.e. a solid dark line appearing to cover around 15-20% of the pixels in a single image row;
2. A “dotted dash” i.e. like the lines above, but with values varying from pixel to pixel;
3. A broad “band” feature covering several image rows, in which there appear to be narrow vertical bands caused by alternating pixel values;
4. Inspection of images displayed with different settings from above reveals an additional “dash” artefact with similar appearance to type (1) on the image display, but different data characteristics.

IDL software has been developed to read a UBT file and plot the content in different ways in order to inspect the data values when there are data artefacts present. The tool checkUBTartefacts.pro reads a UBT file using standard code based on the examples in [AD 2], then displays the 11µm nadir channel image. The 11µm channel is chosen because it is always active and the 12µm contains the blanking pulse indication (pixel values are negated to show presence of the pulse).

checkUBTartefacts.pro displays by default:

1. The 11µm nadir image. The current code uses the tv and hist_equal functions to give the best contrast.
2. The 11µm nadir raw image displayed by the tvscl function, in case different features are revealed.
3. Line plots showing, for each image row, the ratio of the odd and even pixel gains and the individual odd and even gain values. These plots are “flipped” so that features are aligned with the nadir image scans.

Within the IDL session, if the user points their mouse and clicks at a specific point in image window (1), line plots of the selected scan values can be displayed. Plots are made for 3 scan lines in total, 1 scan either side of the selected (to account for possible inaccuracy in clicking). Two plots per scan show:

1. A line plot of pixel values for an entire scan;
2. A zoomed version, covering 80 pixels from the selected scan for a smaller range of pixel values.

The reported generation of an artefact-free version of a scene UBT by processing from a different Level 0 tape suggests that data corruption might be the cause. The ATSR-1 data does not include a CRC check value so it is not possible to confirm this via a checksum.

Approximately 20 affected scenes from UBT products have been checked using the IDL tool to date.

6 Results and possible solutions

Examples of the 11µm nadir image display options are shown in Figure 2 below. On the left is the default display option used for feature selection by the checkUBTartefacts tool, which generally gives good contrast and highlights artefacts. The raw image data displayed by tvscl (centre) can highlight artefact features not always seen in the default view. On the right, the raw image appears speckled and obscures detail.

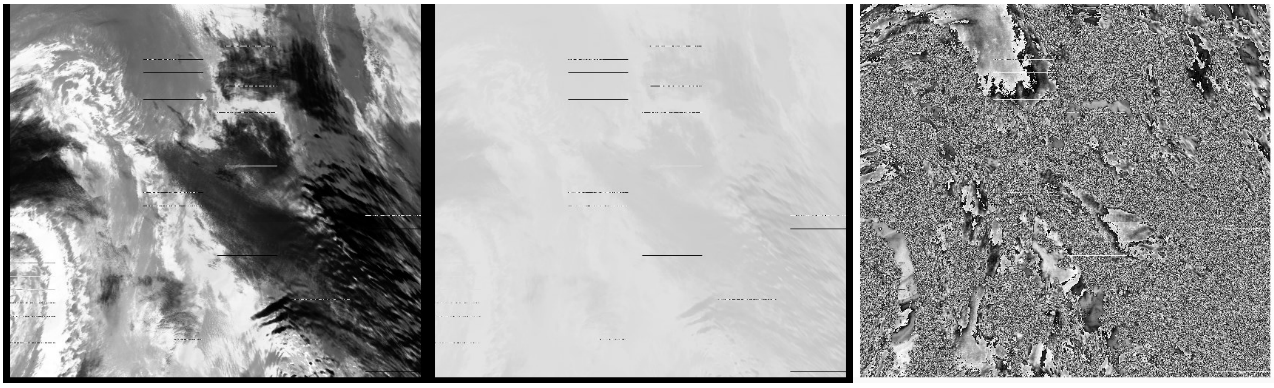

Figure 2: 3 image display options for the 11µm nadir channel. Left to right: histogram-equalised, tvscl (raw image), tv (raw image).

6.1 "Dash" artefact (flagged values)

Inspection of the scan line plots show clearly the presence of the "dash" artefact, as seen in Figure 3. Zooming in or printing pixel values shows that these pixels are set to the flag value -4 (no signal in channel or 0 count). In many of the cases checked, the flagging persists for 80 pixels and affects all channels. Cases are also seen where the extent is 60 pixels, e.g. when the artefact appears at one end of a scan.

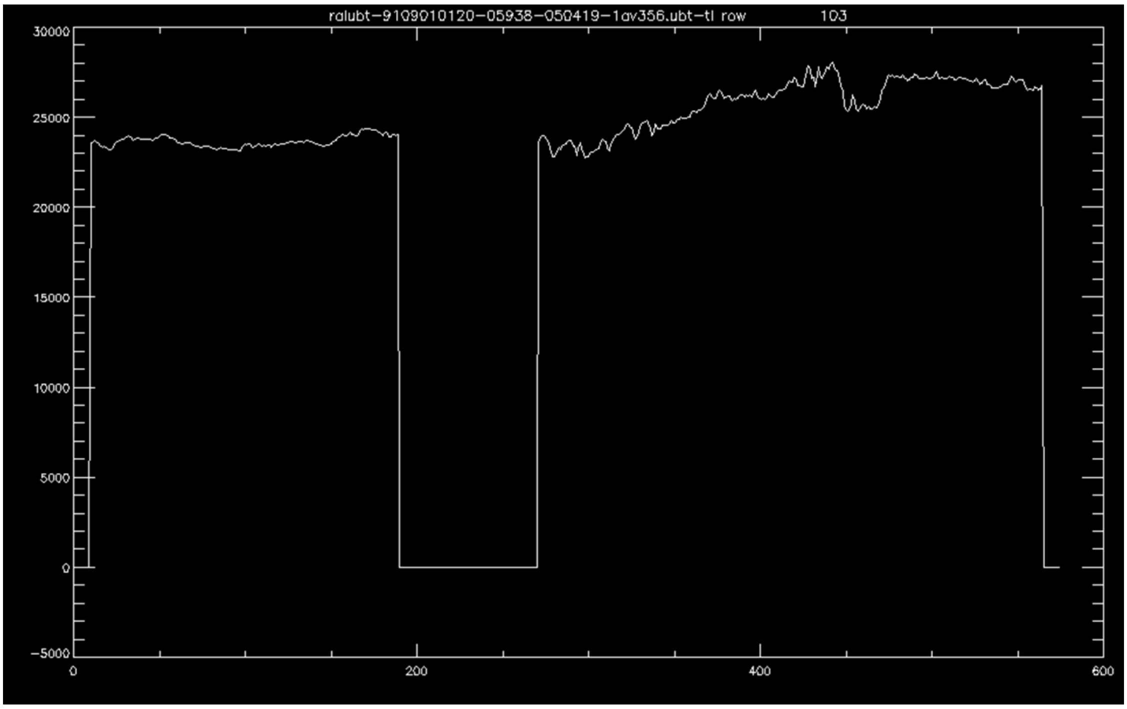

Figure 3: pixel values along an image row containing a "dash" artefact

Modification of the line plot code to include all available channels shows that all channels are affected, as illustrated in Figure 4.

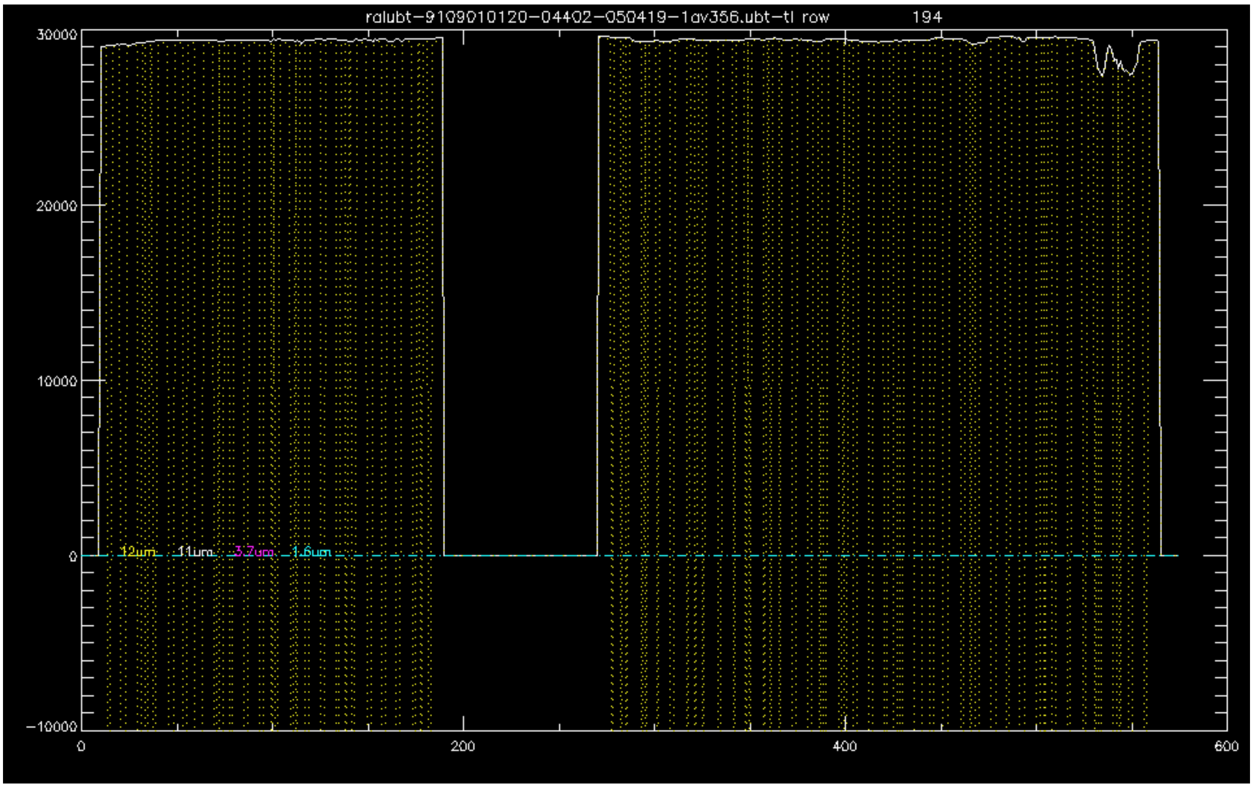

Figure 4: plot along a scan containing a “dash” artefact, showing all channels. The dotted lines alternating from positive to negative values are caused by the negation of 12µm channel values to indicate the presence of the blanking pulse.

There is therefore no need for post-processing to flag these features. However, if reprocessing from level 0 is considered for other reasons, it should be kept in mind that processing from different tape sources or different Level 0 files can generate different results so it may turn out that the missing data values may be available.

6.2 “Dash” artefact (apparently valid values)

Comparison of the two scene images in Figure 5 shows that some “dash” features which appear as solid black lines in the histogram-equalised view appear fainter in the “raw” image view. These features are of a different type in which the pixel values are not flagged and appear to be valid. The values are lower than adjacent pixels and scans, but not so low as to appear obviously invalid, as shown by Figure 6. There is a step change in value relative to the pixels at either end of the feature in the same scan and relative to the adjacent scans.

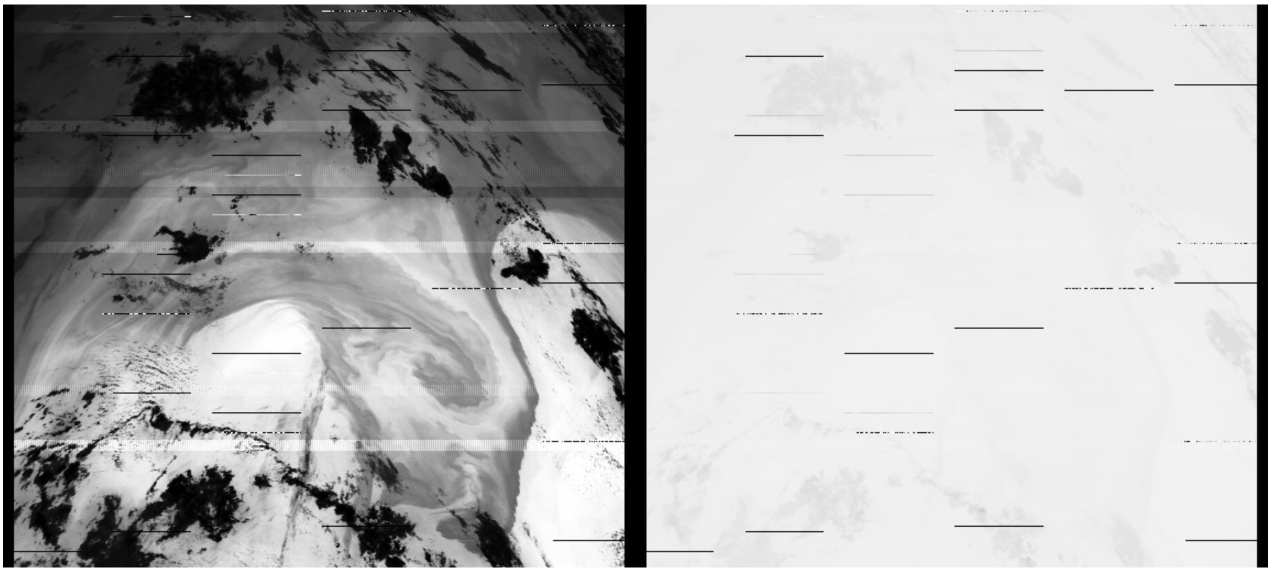

Figure 5: histogram-equalised and raw 11µm channel images for scene ralubt-9109010120-04402

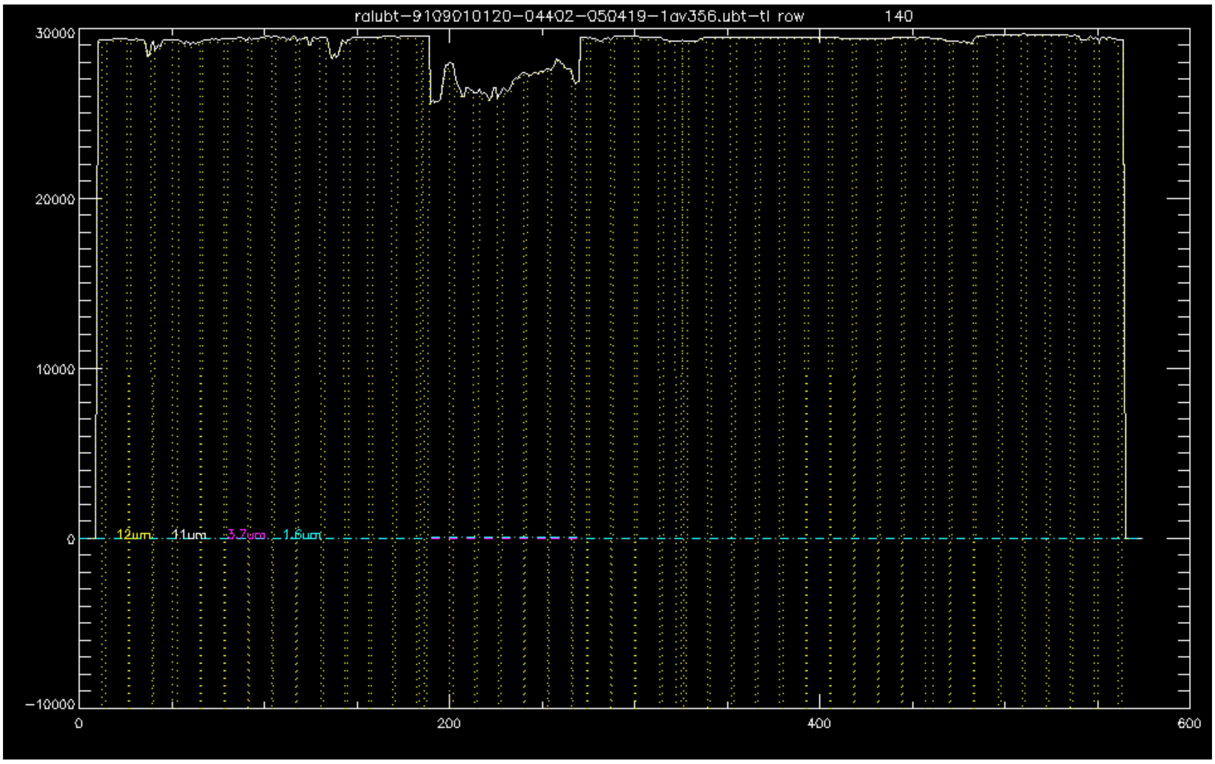

Figure 6: pixel values from scan 140 in scene ralubt-9109010120-04402

This type of artefact is clearly detectable through visual inspection of the nadir image, but does not have an obvious signature in the scan data. It might be possible to automate detection by looking for the step change versus the scans either side, affecting a range of adjacent pixels within a scan line.

6.3 “Dotted dash” artefact

Inspection of image row line plots shows widely varying values when the “dotted” or broken dash features are present, as seen in Figure 7.

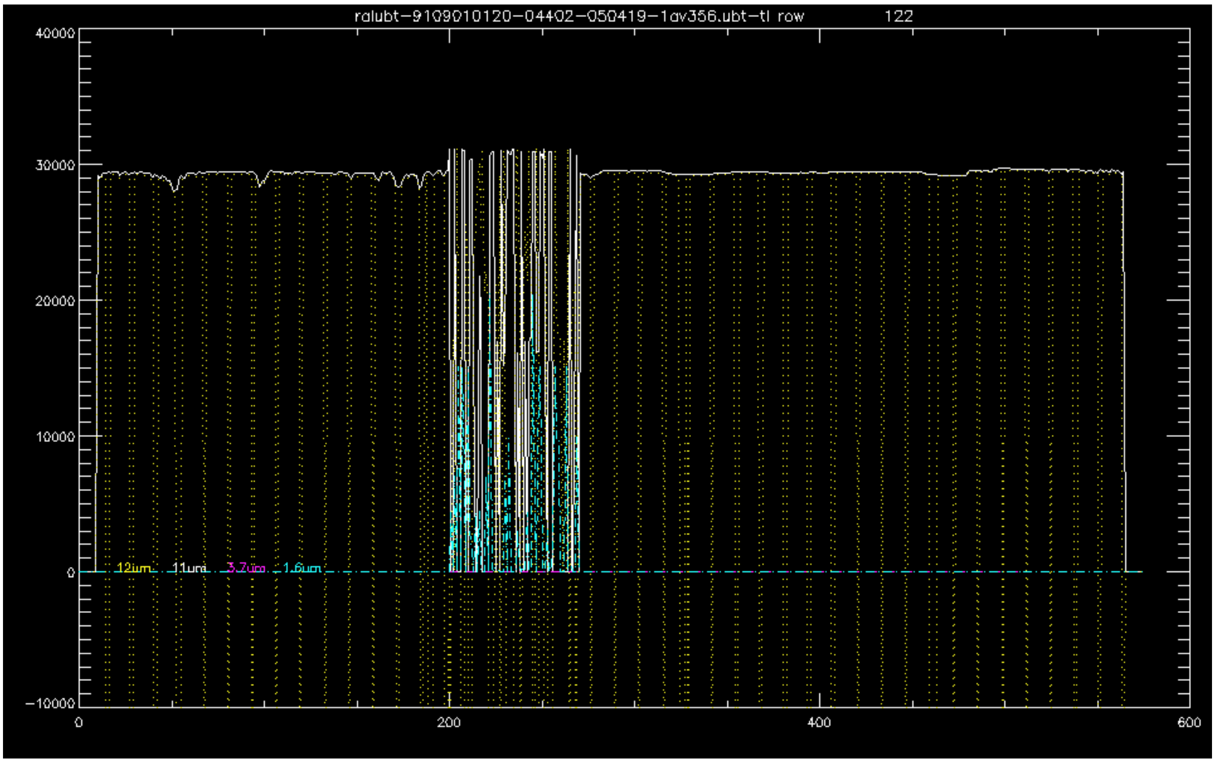

Figure 7: pixel values in a scan line containing a “dotted dash” artefact

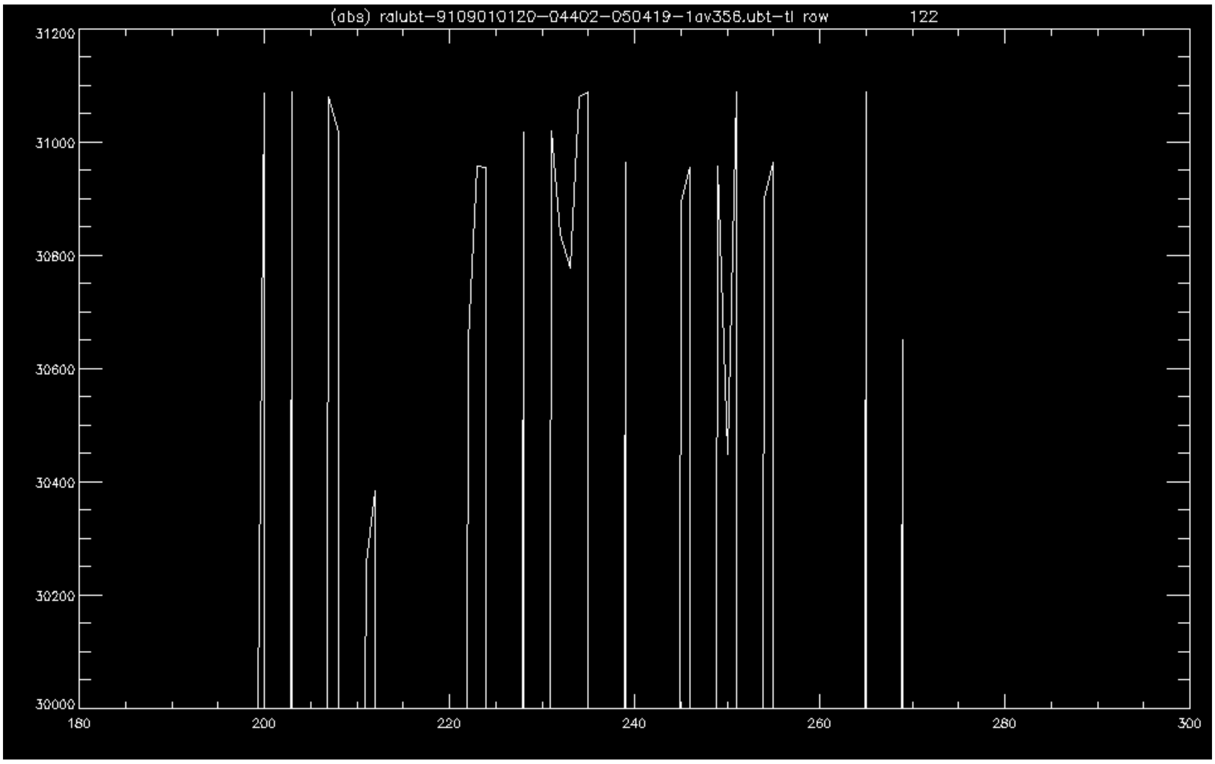

Figure 8: zoomed plot of pixel values from a scan line containing a dotted dash artefact

Within these features some pixels correspond to flag values as in the full dash cases, some have values higher than other pixels in the row, and others have intermediate values which are considerably lower than the surrounding (and assumed valid) pixel values. Again, the artefact typically affects 70 to 80 pixels. In this case it is harder to give an exact number of affected pixels since the values are not easily distinguished. Again, all channels are affected.

It is not immediately obvious whether the maximum values seen are "non-physical" and therefore easily distinguishable from valid pixel values for creation of a detection and flagging algorithm.

An option for identifying this type of artefact might be:

1) to look for ranges of, e.g. 50 pixels in a row in which more than a threshold percentage of pixels are set to the -4 or -5 flag values.
2) To look at the standard deviation over a similar range of pixels, as an indication of wide variation.

However scanning an entire image row by row in this way could be quite expensive in processor time.

6.4 "Band" artefact

As seen in Figure 1, the "band" artefacts affect several scans and extend across the entire scan. Line plots along the scan line reveal more information about the pixel to pixel variation.

Even at the full-scale resolution in Figure 9 it is possible to see a distinct zig-zag from pixel to pixel.

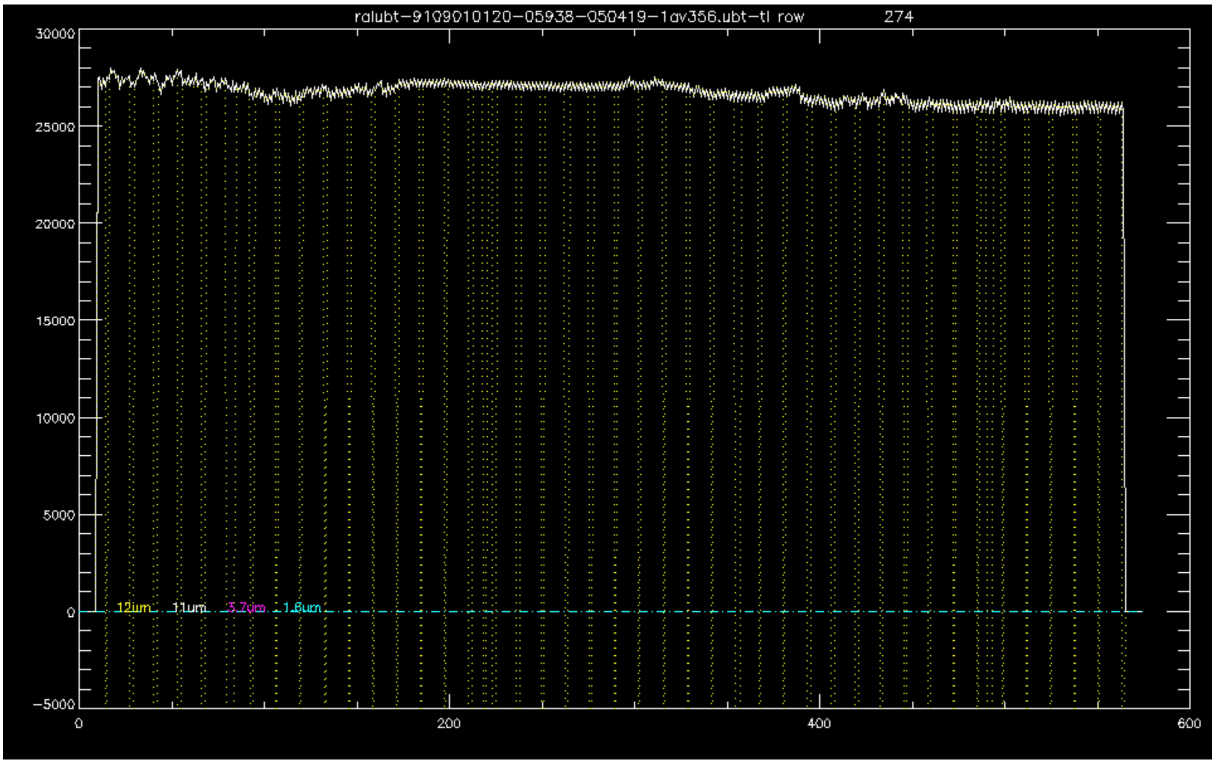

Figure 9: a line plot along an image row affect by a "band" artefact

Zooming in, as in Figure 10, shows a distinct pattern in the values between adjacent pixels. This type of variation could not be distinguished in earlier investigations using plots of 12µm channel data due to the presence of the blanking pulse values.

Superposition of extra lines representing only the odd or even pixel values (the blue and red lines in Figure 10) shows that there are clear differences between the two pixel sub-sets.

Within the UBT product the only data available separately from odd and even pixels are the calibration gains and offsets, which are stored for each detector record, i.e. each scan. Separate values are stored due to the use of two integrators, because the integrator response time is too slow to use the same one for all pixels.

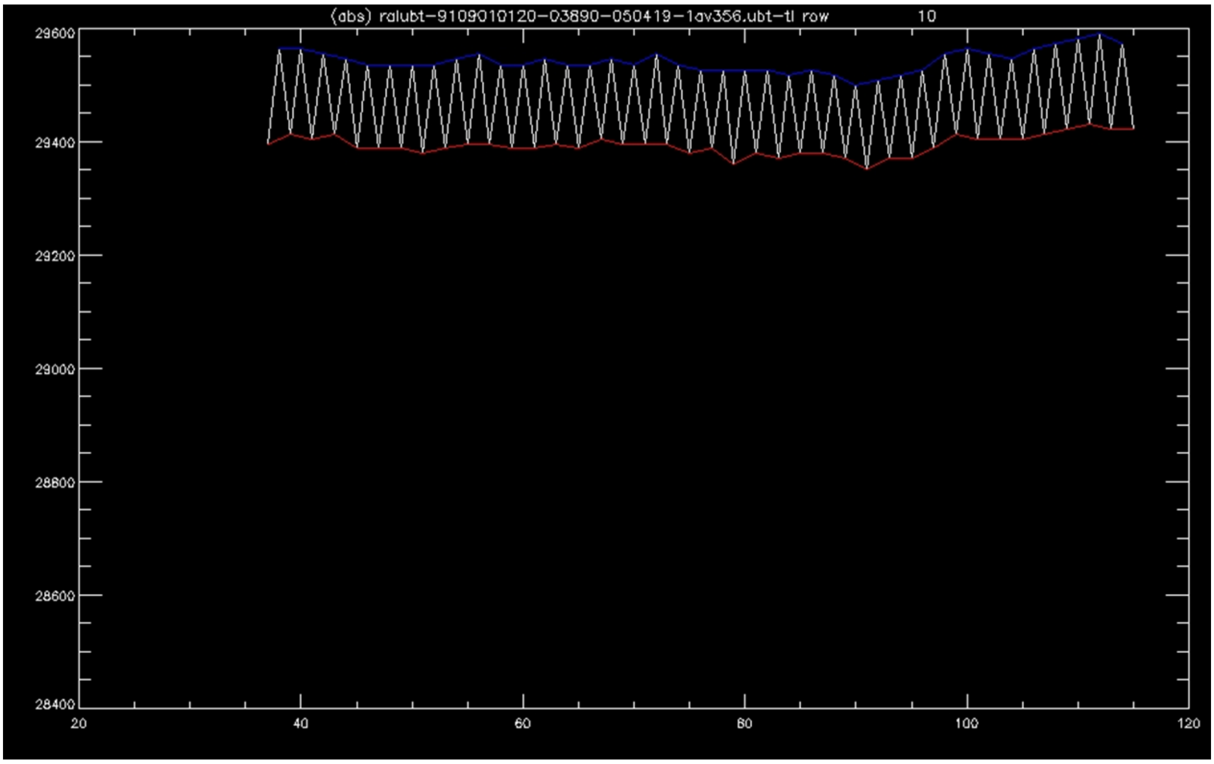

Figure 10: zoomed line plot of a UBT image row affected by a “band” artefact. The blue and red lines show only the even and odd pixel values.

Figure 11 includes a line plot of the odd/even gain ratio and the individual gains, with x and y-axes flipped and scaled to so that features line up with those in the adjacent channel image. There is a clear correlation between the band artefacts in the image and those scans where the gain ratio deviates from approximately 1.0. It was also found from inspection of this plot that there is a band feature which is not immediately visible on inspection of the image, but whose presence is indicated by the gain ratio plot.

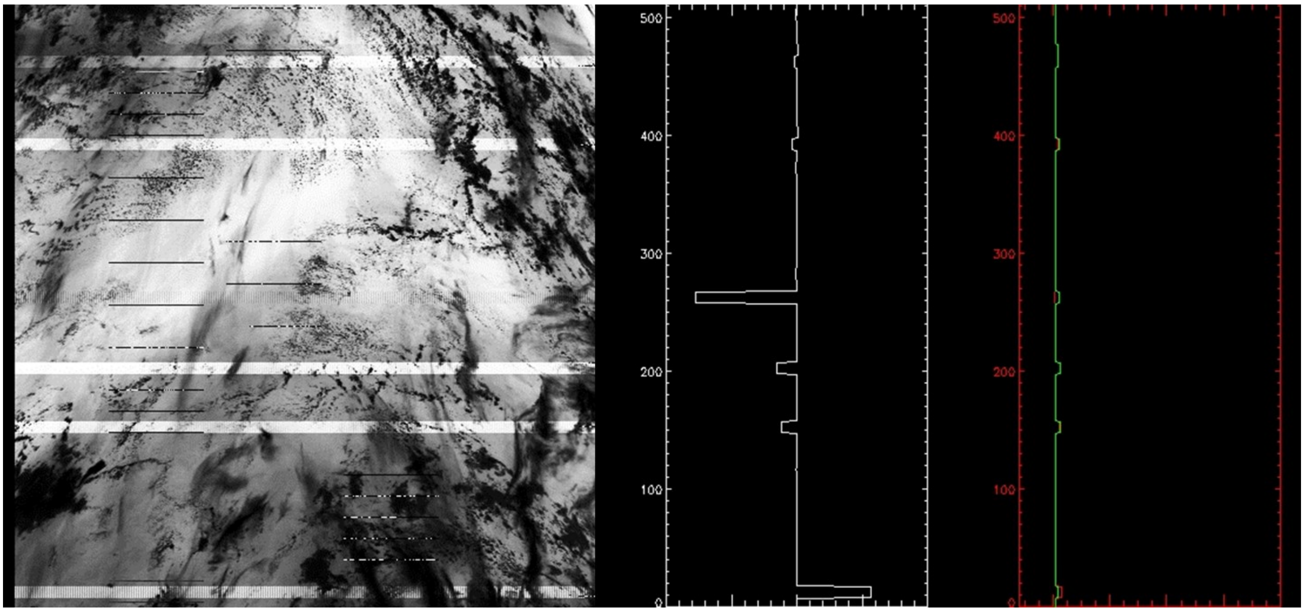

Figure 11: UBT channel image containing band artefacts, with line plots of the odd/even gain ratio in each row and the individual odd/even gain values

An experimental flagging algorithm was implemented in the checkUBTartefacts tool, flagging out all rows where the odd/even gain ratio differed from 1.0 by greater than +/-1%. This algorithm successfully removed the bands from one example image as shown in Figure 12 . However, a second case was found which contained additional horizontal bands without the vertical striping (Figure 13).

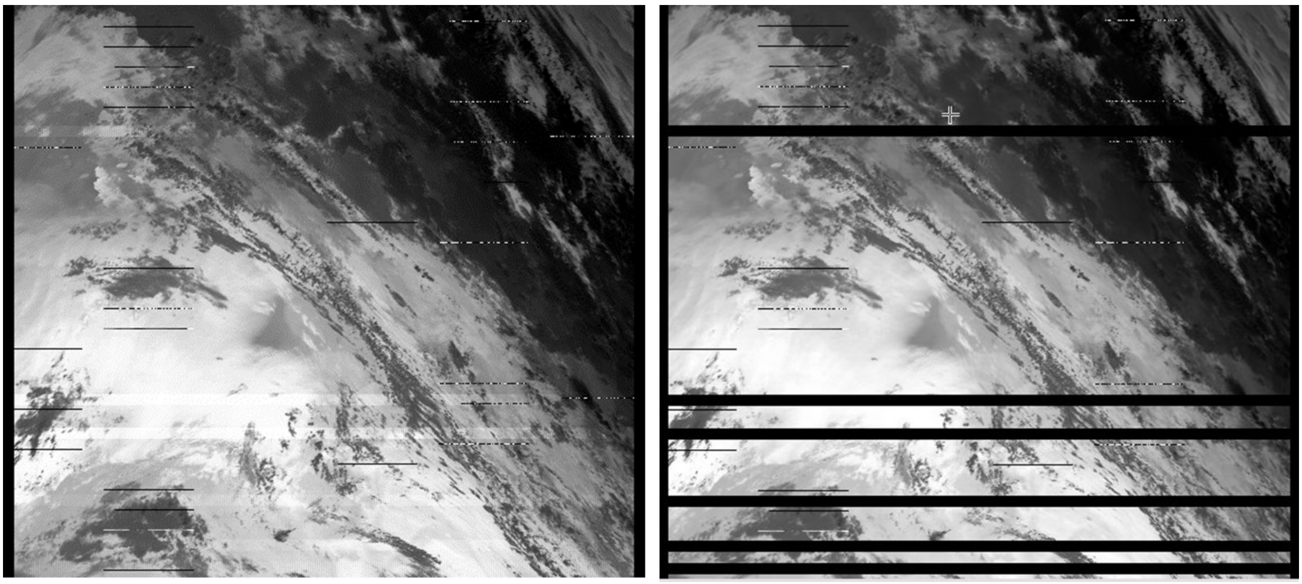

Figure 12: a UBT image affected by band features (left) and the result of an experimental pixel flagging based on the odd-even gain ratio (right)

Further investigation is needed to determine whether the non-flagged horizontal band case is genuinely a different effect, or whether the flagging threshold for the odd-even gain ratio needs refinement.

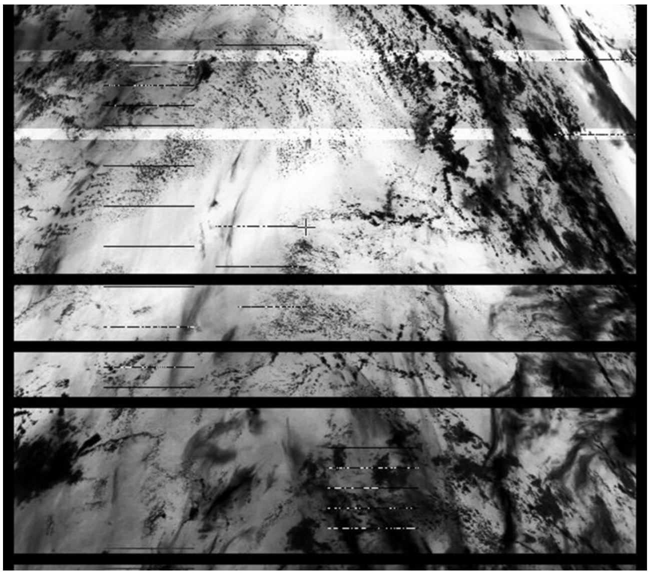

Figure 13: a UBT scene in which the attempted flagging based on odd-even gain ratio did not remove all band features.

6.5 Processing and flagging options

The processing / flagging options available are (1) to reprocess from Level 0 to UBT or (2) to post-process the UBT archive adding flags to affected scans or pixels.

6.5.1 Scanning the ATSR-1 archive

Both processing approaches require a scanning tool to be run on the entire existing ATSR-1 UBT archive. In advance of reprocessing from Level 0, an archive scan is needed to generate a list of target products or periods for reprocessing. If the post-processing approach is preferred, the archive must be scanned either as part of the flagging, or as a pre-processing to generate a list of products to be flagged.

Scanning the UBT archive is made easier by the CEMS / Jasmin system, in which the CEDA data directories are directly accessible. The ATSR-1 UBT products are gzipped when archived, so the procedure for scanning would be, for example:

1. Create a working directory.
2. Copy and unzip a day's UBT files into the working directory.
3. Scan each unzipped file.

6.5.2 Reprocessing from Level 0 to UBT

There are two possibilities for reprocessing from Level 0:

1. Reprocess using the current SUPPLE code and check whether the resulting UBT products are free of artefacts.
2. Build in artefact detection and flagging into the SUPPLE code.

As noted earlier, reprocessing from a different Level 0 source file from the archived UBT product gives the possibility of generating a new UBT product free of artefacts. Unfortunately there is no traceability information in the UBT product header to indicate the original Level 0 source file or tape, or the receiving station. Level 0 files include a station identifier in the main product header.

The recent re-transcription of Matera receiving station data to Level 0 files by the DSI/X-PReSS project provides a new Level 0 archive which is available online. Previously any reprocessing from Level 0 to UBT would have had to use tape sources and would be impractical. The Matera re-transcription is currently understood by RAL to be based on an accumulation of tapes from different receiving stations, rather than data collected only by Matera.

Access to the Matera data is currently via the CEMS ftp server, which makes automated access for bulk production processing difficult. The data will soon be archived by ESA and removed from the CEMS server. Bulk transfer from the ESA archive could take a considerable time. Processing would be easier and more efficient if a data archive was easily accessible from the processing environment, as with the UBT archive on the Jasmin system.

The RAL SUPPLE Level 0 to UBT processor has recently been extended to handle ATSR-1 packets. This processor is a port to Linux of the SADIST-2 processor, performed in 2008. The SUPPLE processor has limitations. Only the most common (ATSR-2) pixel maps can be processed as these were the priority for the original porting work. The ATSR-1 data set is thought to contain data with a wider range of pixel maps which were used during commissioning to determine the most effective way to operate the instrument. Extending the SUPPLE code to process additional pixel maps and the associated data compression modes is likely to be a long task.

Test processing of a small number of DSI Level 0 files has been done using SUPPLE, but so far this has been unsuccessful or inconclusive. The Level 0 input files were chosen to match times when artefacts were known to be present in the UBT archive. However the time range of the resulting UBT files did not extend to the points at which artefacts were found. It is possible that the data includes periods where the instrument was operated using pixel maps and data compression modes that cannot be processed by SUPPLE, as described above. It is also possible that ATSR data was not present for the entire period indicated by the Level 0 file name. This may occur because Level 0 files contain data from at least one other instrument.

Approach (1) is a possibility, but without knowledge of the original Level 0 source files, and without successful test processing runs, results are not guaranteed. The approach could be described as “process and hope for the best”.

Approach (2) is probably more complex and less practical than flagging existing UBT products via an IDL post-processor and is not discussed further.

6.5.3 Post-processing of UBT files

Post-processing of affected UBT files and addition of new flags may be a better option than reprocessing from Level 0 because the UBT archive is directly accessible via CEDA’s Jasmin system.

Flagging algorithms used in post-processing need careful validation over a wide range of scenes in order to prevent false-positive flags. This could be very time-consuming given the nature of some of the artefacts.

Post-processing would also require a scan of the entire archive since there is no comprehensive listing of affected periods or data products. Either (1) the archive could be scanned in advance, to identify a list of products for flagging, or (2) to avoid duplication of work, the scan and flagging stages could be combined. The flagging tool should be written such that a new UBT file is generated only if artefacts are found.

Flag values to be assigned to the different types of artefact are TBD. The current flag values are described in the following section.

Also TBD is how to distinguish a new, flagged UBT file from the original file. An indication in the filename would be useful. The version number in the product filename is a SADIST software version and should not be changed, since the basic processing algorithm and calibration that generated the pixel values will not change. The filename contains a processing date, which could be updated, but this alone is not sufficient to distinguish a flagged file from a newly-processed file using the same software version. If there are no restrictions on extending the length of archived UBT filenames, an addition of an ‘f’ to the version string or the filename suffix might be possible. However, changing the filename length may risk causing problems with existing data handling tools or the database of UBT products maintained by NEODC.

Flagged files then need to be merged with the existing archive, replacing the original files in the data set made available to users. Replacement is at the individual file level rather than all files over a range of days, but should be reasonably easy to automate based on the unique combination of orbit time and along-track distance in the filename.

6.5.4 Data flagging

The following flag values are set by the SADIST/SUPPLE code and listed in [AD 2]:

-1 Entire scan absent from telemetry
-2 Pixel absent from telemetry (possible reasons are disablement of channel or visible channel fixed to narrow swath)
-3 Pixel not decompressed, due to error reading packet validation
-4 No signal in channel (zero count)
-5 Saturation in channel (maximum count)
-6 Derived radiance outside range of calibration
-7 Calibration parameters unavailable for pixel

The SUPPLE code also defines -8 as pixel unfilled and MAX_SINGLE_PIXEL_ERROR as 8, so there are no currently “spare” values. However, the brightness temperatures and reflectances are stored as (value * 100) so adding a -9 flag does not create any significant risk of confusion with a genuine data value.

7 Conclusions

It was decided at the QWG telecon on 5th May 2016 that the current work package should be concluded. The current understanding of artefacts and potential further work required are summarised below.

7.1.1 Characterisation / identification of artefacts

The “dash” artefact containing flag values need not be considered for post-processing. Reprocessing from Level 0 gives the possibility that UBT products free of these artefacts can be generated but it is not known whether the Level 0 files now available differ from the originals.

The “dash” artefact containing apparently valid values is not well characterised and there is no current detection algorithm. Further work is needed to develop an algorithm, possibly based on checking adjacent scans. Advice from the science team is needed to help understand valid and invalid pixel values.

The “dotted dash” artefact needs further work to develop an identification algorithm. The standard deviation along part of a scan may be an indicator, but a suitable threshold value must be identified.

The “band” artefact can be identified by the odd/even pixel gain ratio. Further work is needed to check a larger number of affected scenes and to identify a suitable threshold value.

7.1.2 Software development

Post-processing and flagging of affected UBT files is considered the best approach.

Decisions are needed on (1) suitable flag values for affected pixels and (2) identification of flagged UBT files through the filename (or otherwise).

The detection algorithms can then be built into an IDL tool. In all cases identified so far, each artefact affects all channels, so checks need only be performed on one channel. The 11µm is suggested as it is always present and unaffected by the blanking pulse flag.

A control layer is then needed to handle the data extraction from CEMS, unzipping, removal of inputs after processing and caching of new files for archiving.

7.1.3 Algorithm validation

Detection and flagging algorithms, once developed, need thorough validation based on many different scenes in order to eliminate false positives and ensure correct implementation.

7.1.4 Bulk processing and archiving

Bulk processing of the entire ATSR-1 archive (1991-1997) is expected to be reasonably possible to automate, reducing the total effort required, but will take some time to run.

Quality checks on the flagged UBT products are TBD. No Quality Checking (QC) tools exist for UBT products, other than display tools which could be used for visual inspection, which may require a lot of operator effort depending on the proportion of products checked. The effort required depends to some extent on the level of confidence in the validated processor.

Archiving should be reasonably possible to automate, for reasons described earlier, but will be delayed until after QC. Assuming only a small proportion of products are affected, the archiving step should be relatively quick and should not require a large amount of new storage space.